

METHOD, SYSTEM, AND PROGRAM FOR MANAGING A SPEED AT WHICH  
DATA IS TRANSMITTED BETWEEN NETWORK ADAPTORS

BACKGROUND OF THE INVENTION

5 1. Field of the Invention

[0001] The present invention relates to a method, system, and program for managing a speed at which data is transmitted between network adaptors.

2. Description of the Related Art

- 10 [0002] In a network environment, a network adaptor card on a host computer, such as an Ethernet card, Fibre Channel card, etc., will receive Input/Output (I/O) requests or responses to I/O requests initiated from the host. Often, the host computer operating system includes a device driver to communicate with the network adaptor, and provide an interface between the host operating system and the network adaptor.
- 15 [0003] When two Ethernet cards communicate over a network, they both engage in an auto-negotiation protocol to select a highest possible common mode speed at which both are capable of communicating, such as 10 megabits per second (Mbps), 100 Mbs, 1000 Mbs, half duplex, full duplex, etc. During auto-negotiation, the Ethernet devices transmit a series of link pulses that advertise their capabilities to the other device that
- 20 shares the link segment. Devices transmit a Link Code Word (LCW) that is encoded with the device transmission capabilities. The hardware in the Ethernet cards is configured to select the highest common speed between the two devices based on the capabilities communicated in the link pulses. Further details of the auto-negotiation process to select a highest common speed between two network devices is described in the publication
- 25 IEEE Standard 802.3-2002 (Copyright by the Institute of Electrical and Electronic Engineers (IEEE), 2002), which publication is incorporated herein by reference in its entirety.

[0004] With prior art Ethernet auto-negotiation, the two linked devices communicate at the highest common speed, regardless of whether the highest speed is needed.

Transmissions at higher speeds consume greater power than lower speed transmissions.

Moreover, devices may be consuming additional power to transmit at the highest

5 common link speed when such high speeds are not needed. For instance, lower speed transmissions may be sufficient when checking electronic mail or "surfing" the Internet. Power conservation may be especially important for battery powered computing devices, such as laptops, hand held computers, etc., and computers deployed in an energy conservation sensitive environment.

10 [0005] For these reasons, there is a need in the art to provide improved techniques for managing the transmission speed between network devices.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0006] Referring now to the drawings in which like reference numbers represent  
15 corresponding parts throughout:

FIG. 1 illustrates a computing environment in which aspects of the invention are implemented;

FIG. 2 illustrates a format of network packets used in accordance with embodiments of the invention;

20 FIG. 3 illustrates data for a transmission change request in accordance with embodiments of the invention; and

FIGs. 4, 5, and 6 illustrate logic to change a transmission speed between network adaptors in accordance with embodiments of the invention.

#### 25 DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0007] In the following description, reference is made to the accompanying drawings which form a part hereof and which illustrate several embodiments of the present invention. It is understood that other embodiments may be utilized and structural and

operational changes may be made without departing from the scope of the present invention.

[0008] FIG. 1 illustrates a computing environment in which aspects of the invention may be implemented. Two computers 2a, 2b include a central processing unit (CPU) 4a, 4b, a memory 6a, 6b, an operating system 8a, 8b, and a network adaptor 10a, 10b. The network adaptors 10a, 10b may be implemented in a card inserted in a slot in the computers 2a, 2b or be implemented on the computer 2a, 2b motherboard, e.g., LAN On Motherboard (LOM). The network adaptor 10a, 10b may communicate with the computer 2a, 2b, respectively, using any bus protocol known in the art, such as Peripheral Component Interconnect (PCI), PCI-X, Communication Streaming Architecture (CSA), PCI-E, etc. The computers 2a, 2b may comprise any computer system known in the art, such as a mainframe, server, personal computer, storage controller, workstation, laptop, handheld computer, etc. In further embodiments, the network adaptor 10a, 10b may be included in alternative types of computing devices, such as telephony device, network appliance, virtualization devices, etc. Any CPU 4a, 4b and operating system 8a, 8b known in the art may be used. The network adaptor 10a, 10b includes a network protocol 12a, 12b implementing the network adaptor functions, such as auto-negotiation and others, to send and receive network packets between the network adaptors 10a, 10b. The network adaptors 10a, 10b further include registers 14a, 14b that maintain information used during adaptor operations, such as the advertised capabilities of the network adaptor, e.g., 10 Mbps, 100 Mbps, 1000Mbps, half-duplex, full-duplex, etc. The network 16 may comprise a Local Area Network (LAN), the Internet, a Wide Area Network (WAN), Storage Area Network (SAN), etc. In certain embodiments, the network adaptors 10a, 10b and network protocol 12a, 12b may implement the Ethernet protocol, token ring protocol, Fibre Channel protocol, Infiniband, Serial Advanced Technology Attachment (SATA), parallel SCSI, serial attached SCSI cable, etc., or any other network communication protocol known in the art.

[0009] The computers 2a, 2b further include a device driver 18a, 18b that executes in memory 6a, 6b and includes network adaptor 10a, 10b specific interfaces to communicate with and control the network adaptor 10a, 10b and interface between the operating system 8a, 8b and the network adaptor 10a, 10b. The computers 2a, 2b may further  
5 include a transport protocol driver (not shown) executing in memory that processes the content of messages included in the packets received at the network adaptor 12a, 12b that are wrapped in a transport layer, such as TCP and/or IP, Internet Small Computer System Interface (iSCSI), Fibre Channel SCSI, parallel SCSI transport, or any other transport layer protocol known in the art. In further embodiments, the transport protocol driver  
10 operations may be performed within the network adaptor 10a, 10b hardware, such as the case with a Transport Offload Engine (TOE) adaptor.

[0010] The computers 2a, 2b also include one or more application programs 20a, 20b that execute in memory 6a, 6b to perform various operations, including transmitting and receiving data through the network adaptors 10a, 10b.

15 [0011] FIG. 2 illustrates a format of network packets 50 that the network adaptors 10a, 10b may transmit. The network packets 50 include an Ethernet header 52, an Ethernet data frame 54, and an Ethernet cycle redundancy check (CRC) code 56 used to perform link-level error detection at any point from start to end and including middle points in the data transmission path between the network adaptors 10a, 10b. The Ethernet packet 50  
20 may include additional fields and flags not shown that are known in the art and used for Ethernet communication. In certain embodiments, certain data bits in an Ethernet packet 50 may be used to signal a network adaptor 10a, 10b to change the transmission speed.

[0012] FIG. 3 illustrates speed change request codes 70 that may be included in the Ethernet packet 50 to indicate a speed change, including a transmission change request 72  
25 that is the code signaling the linked network adaptor to change speeds and the new transmission speed 74. The speed change request codes 70 may be included in the Ethernet packet, in the preamble transmitted at the beginning of the packet or idle transmissions that occur between Ethernet packets. Preambles are typically transmitted

during idle time periods at the beginning of each Ethernet packet transmission and are used for clock synchronization, whereas idle transmissions occur between packets..

[0013] FIGs. 4 and 5 illustrate operations performed by the network protocol 12a, 12b logic of the network adaptors 10a, 10b to effect a change in transmission speed in response to the occurrence of a transmission speed change event. The network protocol 12a, 12b may be implemented as hardware logic in the network adaptor 10a, 10b. A transmission speed change event can be triggered by the device driver 18a, 18b or some other component in the computer 2a, 2b. For instance, the device driver 18a, 18b may implement an application programming interface (API) call that allows the application program 20a, 20b to change the transmission speed. If the application program 20a, 20b anticipates performing operations consuming significant network 16 resources, e.g., such as a network request for large files, such as a multi-media file (e.g., video, audio, etc.), new program or program update to download, etc., then the application program 20a, 20b may invoke the speed change API to cause the device driver 18a, 18b to, in response, perform operations to effectuate the transmission speed change operations. Additionally, a device driver 18a, 18b API may be provided to allow an application program 20a, 20b to reduce the transmission speed. In additional embodiments, the device driver 18a, 18b or some other component in the computer 2a, 2b may request the network adaptor 10a, 10b to adjust the transmission speed in response to detecting changes in network traffic at the network adaptor. For instance, if increased or decreased traffic is detected at the network adaptor 10a, 10b, then action may be taken to cause the network protocol 12a, 12b to perform the operations described in FIGs. 4 and 5. In the description in FIGs. 4 and 5, the network adaptor 10a is described as the local device where the speed change is initiated and the network adaptor 10b is described as the linked device that responds to the speed change initiated by the local device 10a. However, the network protocol 12a, 12b enables any of the network adaptors 10a, 10b to function as a local device or linked device.

[0014] With respect to FIG. 4, after initiating an operation to change the speed (at block 100), the local network adaptor 10a generates (at block 102) a speed change request 70 including a new transmission speed 74 (FIG. 3), such as 10BASE-T, 100BASE-TX, 100BASE-T4, 100BASE-T, 1000BASE-X, 1000BASE-T, etc., and transmits (at block 5 104) the speed change request to the linked network adaptor 10b.

[0015] FIG. 5 illustrates operations performed by the network protocol 12b in the linked network adaptor 10b. With respect to FIG. 5, upon receiving (at block 106) the speed change request, which may be implemented in bits in an Ethernet packet, the preamble that is transmitted at the beginning of the packets or the idle transmissions that 10 occur between packets, the linked network adaptor 10b determines (at block 108) whether the requested new transmission speed 74 is supported by the advertised capabilities indicated in the registers 14b of the linked network adaptor 10b. If (at block 108) the linked network adaptor 10b supports the requested new transmission speed 74, then the linked network adaptor 10a sets (at block 110) the configuration registers 14b to 15 implement the requested new transmission speed 74 and returns (at block 112) acknowledgment to the local network adaptor 10a that the speed change is implemented at the linked network adaptor 10b. If (at block 108) the linked network adaptor 10b does not support the requested new transmission speed 74, as indicated in the registers 14b of the linked network adaptor 10b, then negative acknowledgment is returned (at block 114) 20 to the local network adaptor 10a indicating that the linked network adaptor 10b does not support the requested new transmission speed 74.

[0016] With respect to FIG. 4, upon the local network adaptor 10a, 10b receiving (at block 116) the acknowledgment request, if (at block 118) acknowledgment is positive, then the local network adaptor 10a sets (at block 120) the configuration registers 14a to 25 indicate data transmission at the requested new transmission speed. Otherwise, if negative acknowledgment was returned, then the local network adaptor 10a enters (at block 122) an error handling mode. In this error handling mode, the network adaptor 10a may continue to transmit at the previously negotiated speed, which would be the highest

common transmission speed between the local and linked adaptors 10a, 10b.

Alternatively, the local network adaptor 10a may determine the next higher or lower transmission speed the local network adaptor 10a supports, depending upon whether the speed change request was to increase or decrease the transmission speed, respectively,

5 and then perform the logic of FIG. 4 with respect to a next higher or lower transmission speed.

[0017] In still further alternative embodiments, the local network adaptor 10a may be aware of the transmission capabilities of the linked network adaptor 10b, which would be determined when the local and linked network adaptors 10a, 10b initially establish  
10 communication. In such embodiments, the local network adaptor 10a would select the next highest or lowest common transmission speed between the local and linked network adaptors 10a, 10b as the new transmission speed, depending on whether the speed change request is to increase or decrease the transmission speed, respectively. Such embodiments would avoid the situation where the linked network adaptor 10b returns  
15 negative acknowledgment due to the local network adaptor 10a requesting a new transmission speed the linked network adaptor 10b does not support.

[0018] Still further, the speed change request may request that the speed change be to an absolute highest or lowest speed, which would cause the local network adaptor 12a to attempt communication at the very highest or lowest supported speed, respectively.  
20 Additionally, the speed change request may request a speed that is some number of levels higher or lower than the current speed, assuming the local device supports multiple different speed levels. In such case, the local network adaptor would attempt to establish communication at the requested number of levels higher or lower than the current transmission speed.

25 [0019] FIG. 6 illustrates logic implemented in the device driver 18a, 18b to initiate a speed change between the local network adaptor 10a, 10b and a linked network adaptor. The device drivers 18a, 18b may initiate an operation to change the speed in response to a request from an application 20a, 20b or in response to detecting a change in network

traffic at the network adaptor 10a, 10b, as discussed above. For purposes of this description, device driver 18a will be described as the local device driver 18a in the local network adaptor 10a initiating the speed change operation and device driver 18b will be described as the linked device driver 18b in the linked network adaptor 10b that responds to the speed change request from the local network adaptor 10a. However, both device drivers 18a, 18b can function as the local and linked device drivers 18a, 18b. Upon initiating (at block 150) a speed change operation, the local device driver 18a determines (at block 152) the new transmission speed. In embodiments where the local device driver 18a detects the need to adjust the speed as a result of a predefined change in network traffic, the device driver 18a may include code to determine the appropriate new transmission speed for the detected change. Alternatively, if the application 20a calls an API to change the speed, then either the local device driver 18a, application 20a or the network protocol 12a may determine the new transmission speed.

[0020] If (at block 154) the new transmission speed is slower than the current transmission speed, then the device driver 18a calls (at block 156) the network protocol 12a to update the configuration registers 14a including advertised capabilities to eliminate any settings indicating transmission speeds capabilities greater than the determined transmission speed. Otherwise if the new transmission speed is greater than the current speed, then the device driver 18a calls the network protocol 12a to update (at block 158) the configuration registers 14a including advertised capabilities to eliminate any transmission speeds capabilities less than determined transmission speed. Before updating the transmission speed at blocks 156 or 158, the device driver 18a may reset the registers 14a including the advertised capabilities to the initial default values to undo any previous speed setting changes before the current updating is performed to eliminate advertised speed capabilities above or below the determined speed capability. From block 156 or 158, the device driver calls the network protocol 12a to restart (at block 160) the auto-negotiation process to negotiate with the linked network adaptor 10b to transmit at the highest common speed based on the new advertised capabilities of the



local network adaptor 12a set by the device driver 18a. Thus, if the device driver 18a had eliminated the higher speed transmission in the advertised capabilities registers 14a, then the network adaptors 10a, 10b would establish communication at the highest common speed of the remaining speeds indicated in the capability registers 14a of the  
5 local network adaptor 10a.

[0021] In certain embodiments, the local device driver 18a may be aware of the transmission speed capabilities of the linked network adaptor 10a. In such embodiments, the local device driver 18a may set the new transmission speed to the next highest or lowest common speed depending on whether the speed adjustment is to increase or  
10 decrease the speed, respectively.

[0022] With the described embodiments of FIGs. 4 and 5, changing the network transmission speed can occur without any noticeable delay to the user because the request change is transmitted during normal packet transmissions within a packet or within a preamble, so that the change occurs almost immediately once the linked network adaptor  
15 determines that the requested new speed is supported and returns positive acknowledgment. Thus, the speed change request is transmitted without terminating the linked exchange occurring between the local and linked network devices. Once the positive acknowledgment is returned, both network adaptors may begin transmitting at the requested new transmission speed. Until the positive acknowledgment is returned,  
20 the network adaptors communicate using the previously negotiated transmission speed. The described embodiment of FIG. 6 may take longer to implement the new transmission speed than the embodiment of FIGs. 4 and 5 because the device driver causes the speed to change by restarting the auto-negotiation process to negotiate a speed with the new advertised capabilities of the network adaptor having fewer indicated capabilities. With  
25 the embodiment of FIG. 6, transmissions are delayed until the new transmission speed is auto-negotiated, which may be noticeable to a user. However, with FIG. 6, the speed change request occurs without terminating the linked exchange occurring between the local and linked network devices.

[0023] The described embodiments provide techniques to change the transmission speed used to transmit data between network adaptors. The described embodiments allows a change to a slower transmission speed if current network traffic is relatively low. Lowering the transmission speed can conserve energy consumed by the network adaptor, which may be important for portable devices and energy conservation. Moreover, the described embodiments allow an increase in the transmission speed in the event that traffic has increased at the network adaptor or an application program anticipates a need for additional network transmission speed and bandwidth to accommodate anticipated large data transmissions.

10

Additional Embodiment Details

[0024] The described techniques for changing the data transmission speed between network adaptors may be implemented as a method, apparatus or article of manufacture using standard programming and/or engineering techniques to produce software, firmware, hardware, or any combination thereof. The term "article of manufacture" as used herein refers to code or logic implemented in hardware logic (e.g., an integrated circuit chip, Programmable Gate Array (PGA), Application Specific Integrated Circuit (ASIC), etc.) or a computer readable medium, such as magnetic storage medium (e.g., hard disk drives, floppy disks,, tape, etc.), optical storage (CD-ROMs, optical disks, etc.), volatile and non-volatile memory devices (e.g., EEPROMs, ROMs, PROMs, RAMs, DRAMs, SRAMs, firmware, programmable logic, etc.). Code in the computer readable medium is accessed and executed by a processor. The code in which preferred embodiments are implemented may further be accessible through a transmission media or from a file server over a network. In such cases, the article of manufacture in which the code is implemented may comprise a transmission media, such as a network transmission line, wireless transmission media, signals propagating through space, radio waves, infrared signals, etc. Thus, the "article of manufacture" may comprise the medium in which the code is embodied. Additionally, the "article of manufacture" may comprise a

25

combination of hardware and software components in which the code is embodied, processed, and executed. Of course, those skilled in the art will recognize that many modifications may be made to this configuration without departing from the scope of the present invention, and that the article of manufacture may comprise any information

5 bearing medium known in the art.

[0025] In the described embodiments, certain operations were described as being performed by the device driver 10a, 10b and network protocol 12a, 12b. In alternative embodiments, operations described as performed by the device driver 10a, 10b may be performed by the network protocol 12a, 12b, and vice versa.

10 [0026] In described embodiments, the speed adjustment to a higher or lower transmission speed may include going to a higher or lower megabits per second rate, going from either full duplex to half duplex, or vice versa, or changing to any transmission mode known in the art that is supported by the network adaptors and provides a higher or lower data transmission rate than the current transmission rate.

15 [0027] In the described embodiments, the network protocol comprised the Ethernet protocol. In alternative embodiments, the network protocol may comprise any network protocol known in the art, including Fibre Channel, Wireless Fidelity (WiFi), etc.

[0028] In the described embodiments, the network adaptors are shown as being included in different devices or enclosures. In alternative embodiments, the network  
20 adaptors may be included in a same device or enclosure.

[0029] In certain implementations, the device driver and network adaptor embodiments may be included in a computer system including a storage controller, such as a SCSI, Integrated Drive Electronics (IDE), Redundant Array of Independent Disk (RAID), etc., controller, that manages access to a non-volatile storage device, such as a magnetic disk  
25 drive, tape media, optical disk, etc. In alternative implementations, the network adaptor embodiments may be included in a system that does not include a storage controller, such as certain hubs and switches.

[0030] In certain implementations, the device driver and network adaptor embodiments may be implemented in a computer system including a video controller to render information to display on a monitor coupled to the computer system including the device driver and network adaptor, such as a computer system comprising a desktop, workstation, server, mainframe, laptop, handheld computer, etc. Alternatively, the network adaptor and device driver embodiments may be implemented in a computing device that does not include a video controller, such as a switch, router, etc.

[0031] In certain implementations, the network adaptor may be configured to transmit data across a cable connected to a port on the network adaptor. Alternatively, the network adaptor embodiments may be configured to transmit data over a wireless network or connection, such as wireless LAN, Bluetooth, Wireless Fidelity (Wi-Fi), etc.

[0032] The illustrated logic of FIGs. 4, 5, and 6 show certain events occurring in a certain order. In alternative embodiments, certain operations may be performed in a different order, modified or removed. Moreover, steps may be added to the above described logic and still conform to the described embodiments. Further, operations described herein may occur sequentially or certain operations may be processed in parallel. Yet further, operations may be performed by a single processing unit or by distributed processing units.

[0033] The foregoing description of various embodiments of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above teaching. It is intended that the scope of the invention be limited not by this detailed description, but rather by the claims appended hereto. The above specification, examples and data provide a complete description of the manufacture and use of the composition of the invention. Since many embodiments of the invention can be made without departing from the spirit and scope of the invention, the invention resides in the claims hereinafter appended.